DOCUMENT RESUME

ED 331 854 TM 016 373

AUTHOR Schumacker, Randall E.; Fluke, Rickey
TITLE Rasch-Based Factor Analysis of Dichotomously-Scored Item Response Data.
PUB DATE Apr 91
NOTE 33p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Analysis; Equations (Mathematics); *Estimation (Mathematics); *Factor Analysis; *Item Response Theory; *Mathematical Models; Probability; *Scores; Simulation; Test Items
IDENTIFIERS *Dichotomous Scoring; Eigenvalues; Pearson Product Moment Correlation; Phi Coefficient; *Rasch Model; Tetrachoric Correlation

ABSTRACT

Three methods of factor analyzing dichotomously scored item performance data were compared using two raw score data sets of 20-item tests, one reflecting normally distributed latent traits and the other reflecting uniformly distributed latent traits. This comparison was accomplished by using phi and tetrachoric correlations among dichotomous data and Pearson product-moment correlations among Rasch probability estimates of the same dichotomous data in factor analysis. A sample size of 324 resulted from 36 persons at each of nine score values. Eigenvalues for the phi, tetrachoric, and Rasch-r correlation matrices derived from each of the data sets were computed. The Rasch approach, as a psychometric measurement model, was chosen because it met the assumption of a linear ability continuum underlying dichotomous item response data. Results illustrate the superiority of the Rasch-based technique for factor analyzing dichotomously scored item response data. Six tables, one figure, and a 34-item list of references are included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED331854

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

RANDALL SCHUMACKER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Rasch-based Factor Analysis of Dichotomously-scored Item Response Data

Randall E. Schumacker

University of North Texas

and

Rickey Fluke

Hughes Training Systems

BEST COPY AVAILABLE

Paper Presented
at
American Educational Research Association
April 6, 1991
Chicago, Illinois

TM016373

ABSTRACT

The problem addressed in the study is a comparison of three methods of factor-analyzing dichotomously-scored item response data. This comparison was accomplished by using phi and tetrachoric correlations among dichotomous data; and Pearson product-moment correlations among Rasch probability estimates of the same dichotomous data in factor analysis. The Rasch approach, as a psychometric measurement model, was chosen because it met the assumption of a linear ability continuum underlying dichotomous item response data. Results indicated the superiority of the Rasch-based technique for factor-analyzing dichotomously-scored item response data.

Rasch-based Factor Analysis

of

Dichotomously-scored Item Response Data

INTRODUCTION

In the behavioral and cognitive sciences, factor analysis often involves test item responses assuming only two values, either zero or one (Lawley, 1943; Mulaik, 1972; Cureton & D'Agostino, 1983; Gerbing, 1989; Goldstein & Wood, 1989). Previous attempts to solve the estimation problem in the factor analysis of such dichotomous data have focused on the selection of an appropriate measure of correlation, r, ϕ, or r_{tet}, as estimates of unobservable correlations in the population. Pearson product-moment correlations (r) however are intended for use with continuous, interval-level data; phi coefficients (ϕ) with dichotomous data; and tetrachoric coefficients (r_{tet}) with dichotomous data that are assumed to have underlying continuity.

(Hinkle, Wiersma, & Jurs, 1988).

Mislevy (1986, pp. 9-10) listed three shortcomings of ϕ:

> (1) Values of ϕ are dependent not only upon the strength of the relationship between variables, but also upon the difference between their mean values. The phi coefficient can attain extreme values of -1 or +1 only when the two correlated variables have equal means.
>
> (2) The expression for ϕ is generally augmented by terms that depend on the skewness of the discrete variables, which in the dichotomous case, is a function of their mean values.
>
> (3) When binary variables are the result of dichotomizing continuous variables, the placement of the cutting point directly affects the value of the expected ϕ coefficients.

The tetrachoric correlation coefficient (r_{tet}) has also been utilized for dichotomous variables possessing underlying continuous normal distributions (Kachigan, 1986, p. 211; Lord, 1980, p. 39). Several authors, including Bock and Lieberman (1970), Crocker and Algina (1986, p. 320), and Jöreskog and Sorböm (1986, p. IV.3) have recommended r_{tet} rather than ϕ for factor analyzing dichotomous data, although Lord (1980, p. 21) asserted that "Jöreskog's maximum likelihood factor analysis and accompanying significance tests are not strictly applicable to tetrachoric correlation matrices".

Usage and interpretation problems however hinder the use of r_{tet}. Lord (1980, p. 21) warned that "tetrachoric correlations cannot usually be strictly justified" when verifying the unidimensionality of a set of test items, because tetrachoric correlations are "inappropriate for non-normal distributions of ability; they are also inappropriate when the item response function is not a normal ogive [and] whenever there is guessing" (ibid., p. 20). Muthén (1989) noted that:

> (1) r_{tet} matrices are not assured positive definiteness, which may indicate a violation of the underlying normality assumption, or which may reflect sampling variability (p. 24).
>
> (2) r_{tet} matrices generally yield extremely inflated chi-square values and underestimated standard errors of estimate, as compared with Pearson r matrices (p. 24).
>
> (3) The assumption of underlying normality is questionable when the mean value of a dichotomous variable departs appreciably from .5 (p. 27).

In addition, Guilford and Fruchter (1973, pp. 300-306) observed that:

> (4) r_{tet} is less reliable than r, since it is at least 50 percent more variable.
>
> (5) Large samples (N = 200 or 300) are required when r_{tet} is used to estimate the degree of correlation in the population, although sample sizes of about 100 can be used to test the null hypothesis of zero population correlation.
>
> (6) r_{tet} should be avoided when extreme skewness is present, as when p or q = .90 or so, because the standard error is very large in such cases.
>
> (7) r_{tet} must be avoided when only one cell in the pq by p'q' matrix is empty, or when one cell exhibits a much smaller frequency than the other three. In general, the distribution should be fairly symmetrical along one matrix diagonal or the other.

Kim and Mueller (1978, pp. 74-75) offered the following general proscriptions against the factor analysis of dichotomous variables:

> "Nothing can justify the use of factor analysis on dichotomous data except a purely heuristic set of criteria. . . . Even in dichotomies, the use of phi's can be justified if factor analysis is used as a means of finding general clusterings of variables and if the underlying correlations among variables are believed to be moderate--say less than .6 or .7. . . . If the researcher's goal is to search for clustering patterns, the use of factor analysis may be appropriate. . . . One way of doing this is to use tetrachoric correlations instead of phi's. This approach is only heuristic because the calculation of tetrachorics can often break down and the correlation matrix may not be Gramian".

A rigorous investigation of dichotomous test data however was pursued by researchers who were interested in much more than "a purely heuristic" search for "general clusterings of variables". Statistical techniques that enabled simultaneous investigations of

three or more variables were used, of which at least one was often unobserved or "latent" (Loehlin, 1987). These techniques were collectively labeled by Bentler (1980) as "latent variable analysis".

Other authors have also recommended various approaches. Christofferson (1975) described an approach to the factor analysis of dichotomized variables based on the distribution of the first- and second-order joint probabilities of item values using a generalized least-squares estimation procedure with r_{tet} matrices. Wise and Tatsuoka (1986) used order analysis, modified to include item proximity information, to identify the dimensionality of dichotomous data. Stage (1988) analyzed a dichotomous criterion variable using stepwise logistic regression in conjunction with LISREL output. The LISREL measurement model produced unstandardized weightings on the observed variables which were used to create latent predictor variate values, similar to "factor scores", for the subsequent logistic regression analysis. Kim, Nie, and Verba (1977) recommended factor-analyzing tetrachoric correlations calculated from threshold values obtained from a 2-parameter IRT model, instead of phi correlations calculated directly from raw dichotomous data. This procedure was advocated on the conditions that (a) the underlying factors can be assumed to have normal distributions, and (b) each observed binary variable can be conceived to be the result of dichotomizing potentially continuous underlying variables. The authors argued that many practical factor analysis problems met these assumptions, since "a

normally distributed continuous variable can exhibit a variety of observed forms", including skewness (p. 59).

Muthén (1984) proposed a structural equation model with a generalized measurement part allowing observed variables to take on dichotomous, ordered categorical (e.g. Likert) and/or continuous values. The model required an assumption of distributional normality for the latent continuous variables presumed to underlie the observed dichotomous or categorical indicators, even though Muthén "does not believe that underlying normality is always the most appropriate specification". Dichotomous variable analysis using Muthén's GLS model involved serious drawbacks: (a) problems inherent in phi and tetrachoric correlation assumptions and interpretations; (b) a model "is identified if and only if its parameters are identified" in all three parts (p. 118); and (c) the use of maximum likelihood estimates of sample threshold values. This procedure was inferior to a response modeling approach, as apparently, Muthén later realized (1989). Takane and de Leeuw (1987) formally proved that the marginal probabilities of dichotomous variate values obtained from the 2-parameter IRT model and from factor analysis were equivalent. They however used a factor-analytic procedure of the type described by Muthén (1984). Muthén (1989) later attempted to solve some of the estimation and specification problems of factor analyzing dichotomous data. He recognized that both sides of the factor model must be continuous, and that a response model was needed for the continuous latent variable, "the binomial distribution of which is described by

probabilities" (p. 21). Muthéns' error lay in choosing a 2-parameter IRT model to derive continuous probability values for x', and in using r_{tet} correlations to measure the associations in x'. In justifying his use of tetrachoric correlations, Muthén (1989) stated:

> Because the [x'] variables are continuous and unlimited, [tetrachorics] are the proper correlations to analyze. It is also well known that the phi coefficients are attenuated relative to the tetrachoric correlations (p. 22).

Muthén neglected to mention that tetrachorics are designed for observed dichotomous variables with underlying assumed continuous normal distributions. If the continuous variables are observable, either directly or as the result of a response model, it is pointless to use tetrachorics when there is a universally accepted measure of association designed specifically for such continuous, interval-level data, the Pearson product-moment correlation.

The search for an adequate direct measure of correlation among dichotomous variables is doomed in the context of factor analysis because there is a fundamental error of specification motivating such research. The factor model expresses a linear relationship, but the regression of a dichotomous variable on a continuous variable is illogical (Muthén, 1989, pp. 20-21). The factor model works only when x_i is continuous.

A synthesis of the research findings indicates specific premises concerning the latent variable analysis of dichotomous item-response data. First, latent variable analysis requires a better measure of association than either phi or tetrachoric

correlations. Second, dichotomous item responses should be converted to continuous interval-level data before they are subjected to latent variable analysis. Third, the inter-relationships among such item data can be measured with Pearson product-moment correlations which are better-suited to latent variable analysis than either phi or tetrachoric correlations.

Classical test theory unfortunately offers no explanation about how examinees at different ability levels perform on test items. The Rasch model, as a latent trait theory, however does permit estimation of the influence of ability on item performance (Crocker & Algina, 1986). The Rasch logistic function provides transformed score values that indicate equal-interval locations along a latent linear ability continuum (Wright & Stone, 1979). The advantages of the Rasch model over other psychometric measurement models are (Wright, 1977):

> (1) The Rasch model's assumption of independent estimation of ability and difficulty yields score calibrations that are both sample-free and test-free.
>
> (2) Parameter estimates in the Rasch model are unbiased, consistent, efficient, and sufficient.
>
> (3) The Rasch model is the only mathematical formulation for the ogive-shaped response curve that allows independent estimation of ability and difficulty.

The problem addressed in the study therefore is a comparison of three methods of factor-analyzing dichotomously-scored item response data. This comparison was accomplished by using phi and tetrachoric correlations among dichotomous data; and Pearson product-moment correlations among Rasch probability estimates of

the same dichotomous data in factor analysis. This methodology involved (a) Rasch procedures to convert dichotomous item response data to continuous, interval-level logit values for each item, (b) computation of Pearson product-moment correlations among item logit values, and then (c) use of Pearson correlations among items in factor analysis. This method was compared to prior approaches using phi and tetrachoric correlations among items in factor analysis.

METHODS AND PROCEDURES

The latent variable representation for factor analysis is represented in equation [1]:

[1] $$\chi = \Lambda_x \xi + \delta,$$

where: χ = a vector of q observed variables x_i

Λ_x = a (q x n) matrix of factor loadings of x_i on ξ

ξ = a vector of n common factors (latent variables)

δ = a vector of q errors in measuring χ; unique factors; residuals.

Since the factors ξ and the residuals δ are assumed to be continuous, interval-level random variables, the observed variables x_i must abide by the same assumptions in order for the factor model to apply (Muthén, 1989, p. 20). It is also assumed in this study that δ is uncorrelated with ξ (Jöreskog & Sörbom, 1986, p. I.6), and that $q > n$ (Long, 1983, p. 22).

Since the dependent variables x_i are observed and the independent common factors ξ are not, the parameters contained in

Λ_x and δ from equation [1] cannot be directly estimated by regressing x_i on ξ, as in regression analysis. Instead, indirect estimation is achieved by decomposing the population covariances among the x_i variables according to the covariance equation [2]:

[2] $$\Sigma = \Lambda_X \Phi \Lambda_X' + \Theta_\delta ,$$

in which:

Σ = (q x q) matrix of covariances among q observed variables x_i

Λ_x = (q x n) loadings of x_i on ξ , as in Equation [1]

Φ = (n x n) matrix of covariances among n common factors ξ

Λ' = (n x q) transpose of Λ_x

Θ_δ = (q x q) matrix of covariances among q unique factors δ

(Long, 1983, pp. 25, 33-34).

If the variables are standardized to mean value zero and unit variance, then Σ , Φ , and Θ consist of population correlations. In this study, all such matrices are assumed to contain correlations in the off-diagonal positions, and ones in the principal diagonals.

From a sample of observed data, a sample correlation matrix $\underline{S}$ is calculated. Then $\underline{S}$ is used to derive an estimated population correlation matrix $\hat{\Sigma}$, such that equation [3]:

[3] $$\hat{\Sigma} = \hat{\Lambda}_X \hat{\Phi} \hat{\Lambda}_X' + \hat{\Theta}_\delta ,$$

the components of which are estimates of the population parameters in equation [2]. The problem of estimation in confirmatory factor analysis is finding values for the elements of $\hat{\Lambda}_x$, $\hat{\Phi}$ and $\hat{\Theta}$ so that the predicted $\hat{\Sigma}$ is as close as possible to the observed $\underline{S}$

(Long, 1983, p. 35). The purpose of estimation is to infer population parameter values that best explain the observed relationships among sample data.

The Rasch analytic procedure uses a dichotomous persons by items response matrix to calculate estimates of person abilities (b_v) and item difficulties (d_i). These values are then used to estimate the probability p that person v will give a correct response ($x_{vi} = 1$) to item i:

$$p\{x_{vi} = 1 \mid b_v, d_i\} = \exp(b_v - d_i)/[1 + \exp(b_v - d_i)]$$

(Wright & Stone, 1979, p. 15). The Rasch model is used to calculate continuous probability values and thereby solve the problems of estimation and specification in the factor analysis of dichotomous data.

The relationship between the Rasch and factor model is

$$x' = \lambda \xi + \delta ,$$

where: $x' = p\{x_{vi} = 1 \mid b_v, d_i\}$

$$= \exp(b_v - d_i)/[1 + \exp(b_v - d_i)],$$

in which x' is continuous and x_{vi} is dichotomous.

Data and Factor Model

Two raw score data sets were generated and analyzed separately. One data set reflected normally-distributed latent traits and the other reflected uniformly-distributed latent traits. Twenty "observed" dichotomous variables x_i were constructed, consisting of two subsets of 10 variables. Each subset was hypothesized to load on one of two factors, F_1 and F_2, and assumed to be mutually orthogonal (uncorrelated). Each x_i had a unique error factor δ_i. A graphical representation of this factor analysis model is shown in Figure 1.

Insert Figure 1 Here

By subjecting the 20 x_i vectors to Rasch analysis, continuous response probability values were obtained. These values were then substituted for the dichotomous values of x_i, and the resulting continuous variable x'_i replaced x_i in the factor model in Figure 1.

The 20 dichotomous x_i vectors were constructed by letting each subset of 10 x_i variables represent a subtest of 10 items with incremental difficulties. Thus, there were 11 possible scores on each subtest, from zero correct to 10 correct. The 10 x_i vectors were arranged in columns with difficulty increasing from left to right, from x_1 to x_{10} and from x_{11} to x_{20}, then the most likely

response pattern for any raw score was a Guttman pattern (Douglas & Wright, 1990). It can be shown that for any raw score attainable by two or more distinct response patterns, the second-most-likely pattern can be reproduced by transposing the Guttman (1,0) sequential response pair into the anti-Guttman pair (0,1) as indicated in Table 1.

Insert Table 1 Here

The np values in Table 1 are the coefficients from the expansion of the symmetrical binomial $(p + q)^n$, in which p is the probability of a correct answer, q is the probability of an incorrect answer ($q = 1 - p$), and the exponent n represents the number of test items, in this case 10. The coefficients np were taken from Pascal's triangle (Ferguson, 1981, p. 91), and are numerically equivalent to the expected relative frequencies of occurrence of the score values r if the scores are distributed normally (ibid., p. 89). Thus, to simulate a normal distribution of ability levels, the np values from Table 1 were used as frequencies for the corresponding score levels r. A more manageable sample size of n = 340 was attained by dividing each np by three. For the uniform ability distribution, each score occurred with equal frequency. A sample size of n = 324 resulted from 36 persons at each of nine score values.

Rasch analysis discards persons with raw scores of 0 % (r = 0) and 100 % (r = 10). Therefore, only scores of r = 1 to r = 9 were

represented in the data sets. Each 10-item subset represented a unidimensional scale on which a latent trait or ability could be measured. It was assumed that a subset score r_k varied directly with ability level b_k; that is, the greater a person's ability level, the more items he or she would answer correctly. Misfitting or unexpected responses are always present in empirical measurement. The data sets contained artificially replicated misfit by letting 2/3 of the response patterns in each score group be Guttman patterns, and the remaining 1/3 be next-most-likely patterns. The normal and uniform distributions are summarized in Table 2.

Insert Table 2 Here

To ensure that the two latent ability factors underlying each 20-item test were uncorrelated with each other, the case numbers, which uniquely identify the hypothetical persons, were arranged sequentially in one 10-item subset and randomly in the other. These assignments were made separately for the normal and for the uniform data sets.

Data Analysis

The generated data sets were analyzed as follows:

(1) The dichotomous data were subjected to Rasch analysis, from which probability values p_v were obtained, expressing the likelihood that a person gives a correct response to an item.

(2) Three correlation matrices were calculated for each data set, which consisted of (a) phi = ϕ and (b) tetrachoric = r_{tet} for the raw dichotomous data, and (c) Pearson = r for the Rasch p_v values.

(3) Each correlation matrix was subjected to confirmatory factor analysis (Long, 1983). Each matrix was presumed to manifest the same underlying factor model, as shown in Figure 1.

The above analyses were evaluated by comparing each on:

(1) The maximum internal correlation $\rho^{(*)}$ (Joe & Mendoza, 1989), which is "a measure of total dependence among a set of variables . . [and] is an upper bound to product moment correlations" (p. 220). The sample estimate of $\rho^{(*)}$ is:

$$r(*) = (\lambda_1 - \lambda_p) / (\lambda_1 + \lambda_p),$$

where λ_1 and λ_p are, respectively, the largest and smallest eigenvalues of the sample correlation matrix (ibid., p. 212). The data sets in the present study were constructed to reflect maximum internal correlations in the hypothetical population equal to $\rho^{(*)} = 1.00$. Therefore, the values of r(*) calculated for the phi, tetrachoric, and Pearson product-moment correlation matrices can be used to compare the three correlation methods, since each r(*) <u>should</u> be equal to unity, minus the effects of contrived measurement error (caused by misfitting response patterns, held constant for all three correlation matrices calculated from a given data set).

(2) Since $\rho^{(*)}$ is also "an upper bound to the product of the two largest factor loadings (ibid., p. 220)", this product was compared to the criterion value of $\rho^{(*)} = 1.00$ and to obtained values of r(*) for each data set.

(3) The proportion of variance in the input variables accounted for by the two common factors F_1 and F_2 is equal to:

$$\Sigma\, h_i^2/L\ ,$$

where $\Sigma\, h_i^2$ is the sum of the squared communalities for all input variables, and L is the number of input variables (Norusis, 1988, p. B-46). This proportion was compared to a hypothesized population value of 1.0 for each data set, as a means of evaluating the factor solutions.

Procedures

SPSS/PC+ (SPSS Inc., 1988) was used to transform the 10-item generated data sets into 20-item sets, each containing two uncorrelated 10-item subsets. The 20-item normal and uniform data sets were then each subjected to separate Rasch analyses, using MSCALE software (Davis & Wright, 1988; Wright & Stone, 1979). MSCALE output included a difficulty calibration for each item, and an ability calibration for each raw score. These calibrations were used to compute probabilities (p_{ni}) in the normal and uniform data sets.

Eigenvalues for the phi, tetrachoric, and Rasch-r correlation matrices derived from each of the data sets were computed. To extract eigenvalues, a principal components method was used, rather than unweighted least-squares, since:

> "In the [SPSS/PC+] principal components solution, all initial communalities are listed as 1's. In all other solutions, the initial estimate of the communality of a variable is the multiple R^2. . . . These initial communalities are used in the estimation algorithm." (Norusis, 1988, pp. 50-51).

For eigenvalue computations, it was important to use 1's as initial communality estimates, because:

> "When unities are placed in the principal diagonal of R [correlation matrix] then usually $\underline{m} = \underline{n}$ [factors or principal components = original variables]. If some numbers less than unities (estimates of communalities) are placed in the diagonal, and the positive semi-definite [Gramian] property of R is preserved, then $\underline{m}$ will usually be less than $\underline{n}$, and all [eigenvalues] will be real and nonnegative. However, the reduced correlation matrix R (i.e., with communality estimates in the diagonal) will not be positive semi-definite in practice, and both positive and negative [eigenvalues] may be expected." (Harman, 1976, p. 141)

LISREL software (Jöreskog & Sorböm, 1986) was used for confirmatory unweighted least-squares factor analysis of the correlation matrices.

RESULTS

Of the three factor-analytic methods, the Rasch-based procedure rendered pattern matrices most similar to the criterion matrix given normal ability distributions (Table 3). A pattern matrix contains regression weights for the common factors and a structure matrix contains correlations between the factors and the observed variables, but both are similar with standardized variables (Kim & Mueller, 1978, p. 84). The Rasch-r pattern matrix contained 20 significant correlations (one-tailed $p < .05$), as specified by the criterion structure; the other two approaches each contained 16. Table 4 indicates the superiority of the Rasch-based method on two of the three criterion measures described earlier.

Insert Tables 3 & 4 Here

The Rasch-r pattern matrix indicated the strongest overall similarity to the criterion matrix, but the tetrachoric method yielded the highest loadings on six items for the uniform ability distributions (Table 5). However, the tetrachoric method failed to produce significant loadings for two items (numbers 1 and 11);

the other methods estimated significant loadings for all 20 items. Table 6 indicates the Rasch-r method's superiority on the $\Sigma h_i^2/L$ criterion. The tetrachoric method produced the highest $r_{max1}r_{max2}$ value, and the r(*) measures were inconclusive.

Insert Tables 5 & 6 Here

CONCLUSIONS

Rasch-based factor analysis is a viable approach when using dichotomously-scored item response data. Results indicate that Pearson correlations of Rasch estimates performed better than those obtained through phi and tetrachoric correlations among dichotomous data. The Rasch procedure was chosen in favor of other psychometric measurement models because only the Rasch model meets the latent trait analysis assumption that a linear ability continuum underlies dichotomous item response data. This new approach should enable researchers to incorporate results from a variety of dichotomous variables into complex latent variable models which can be analyzed by techniques available in LISREL (Jöreskog & Sorböm, 1986).

The results of this study however are qualified by the following irregularities in the findings:

(1) For the $r_{max1}r_{max2}$ criterion measure in uniform distributions, the tetrachoric value exceeded the Rasch-r value. The difference however was relatively small (.9930 vs. .9264).

(2) Exact r(*) values could not be obtained for several of the data sets, because of negative eigenvalues.

Analyzing P-values versus Ability-Difficulty Differences

Since Rasch probability values (p_{vi}) are restricted to a range of zero to unity, it may seem more appropriate to factor-analyze the differences between Rasch-calibrated person abilities and item difficulties ($b_v - d_i$), which is a theoretically unbounded quantity. However, several considerations led to the decision to use p_{vi} values rather than ($b_v - d_i$) differences as input variables in the present study.

First, even though ($b_v - d_i$) is theoretically unbounded, in practice differences outside the range of -2 to +2 occur in only about 1% of observations (Smith, 1988, pp. 660 & 662).

Second, it has been assumed in this study that an observed item response evaluated at either zero or unity represents a dichotomized manifestation of some continuous unobservable variable. The most logical conceptualization of this unobservable variable across all possible varieties of cognitive abilities and objectively-scored measuring instruments is the latent probability of giving a correct response to an item. Since probabilities of

less than zero or greater than unity are not defined, it is unreasonable to expect a quantitative probability variable to stray outside these bounds.

Third, the use of p_{vi} preserves consistent scaling across latent variables. Replacing dichotomous responses with $(b_v - d_i)$ values would produce response measurements differing in scale across latent variables. Such scale variation could have significant substantive effects on the results obtained from covariance-matrix-based applications of a scale-dependent factor analytic method such as ULS (Long, 1983, p. 79).

Fourth, examinees with perfect scores on a latent ability scale (r = 100% or r = 0%) are eliminated from the Rasch analysis of that scale. Using a p_{vi}-based method, these examinees can be included in subsequent factor analysis, because their observed responses of unity or zero are, appropriately, at the extremes of the p_{vi} value range. But there can be no $(b_v - d_i)$ values for examinees with perfect scores, and their original observed responses are incompatible with $(b_v - d_i)$ scaling. Therefore, these examinees would be excluded from all phases of $(b_v - d_i)$ based factor analysis.

Finally, there are precedents in the psychometric literature for replacing dichotomous values with probabilities prior to factor analysis. Muthén (1989, p. 21) argued that:

> "The y* variable may be thought of as the specific tendency to report a certain symptom [y]. . . . The relationship between y and y* leads to a nonlinear relationship between y and [a common factor F], expressing not the value of y but the probability of y as a function of [F]. This is appropriate because what is needed is a response model for a discrete y, the binomial distribution of which is described by probabilities."

On the other hand, the literature contains no examples or arguments in favor of subjecting $(b_v - d_i)$ values to factor analysis. There may, however, be some advantages to a factor-analytic method based on $(b_v - d_i)$, which could be profitably explored in subsequent research.

Despite the benefits expected to be derived from further study of the new Rasch-r procedure, with a variety of data sets and perhaps with fitting functions other than ULS, and despite the limited explanations provided for specific irregularities, the procedure has clear advantages for researchers who use confirmatory factor analysis with dichotomous item response data. Therefore, its use is strongly recommended in such contexts.

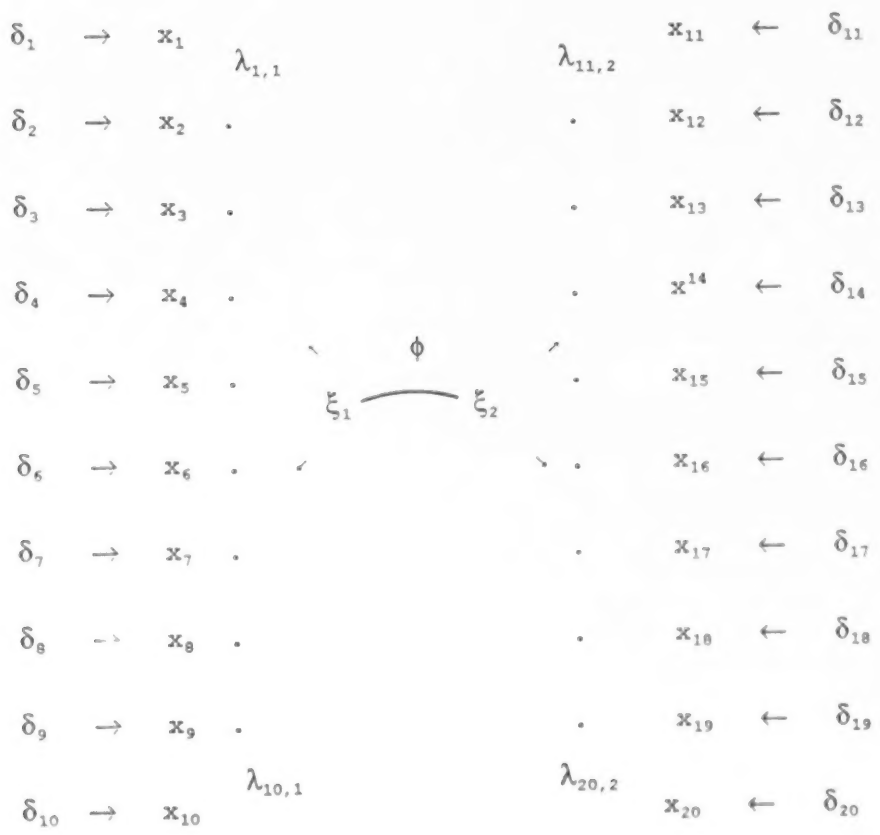

Figure 1: Factor model with 20 observed dichotomous x_i items

---

Legend

δ_i = unique factors; measurement errors in x_i

x_i = observed variables, dichotomously scored (1 = right, 0 = wrong)

λ_{ij} = loadings of variables x_i on common factors

ξ_j = common factors

ϕ_{12} = common factor correlation

Table 1: Probable Dichotomous Response Patterns for 11 Score Values

r	np	Most Likely	2nd-Most Likely
0	1	0 0 0 0 0 0 0 0 0 0 0	N O N E
1	10	1 0 0 0 0 0 0 0 0 0 0	0 1 0 0 0 0 0 0 0 0 0
2	45	1 1 0 0 0 0 0 0 0 0 0	1 0 1 0 0 0 0 0 0 0 0
3	120	1 1 1 0 0 0 0 0 0 0 0	1 1 0 1 0 0 0 0 0 0 0
4	210	1 1 1 1 0 0 0 0 0 0 0	1 1 1 0 1 0 0 0 0 0 0
5	252	1 1 1 1 1 0 0 0 0 0 0	1 1 1 1 0 1 0 0 0 0 0
6	210	1 1 1 1 1 1 0 0 0 0 0	1 1 1 1 1 0 1 0 0 0 0
7	120	1 1 1 1 1 1 1 0 0 0 0	1 1 1 1 1 1 0 1 0 0 0
8	45	1 1 1 1 1 1 1 1 0 0 0	1 1 1 1 1 1 1 0 1 0 0
9	10	1 1 1 1 1 1 1 1 1 0 0	1 1 1 1 1 1 1 1 0 1 0
10	1	1 1 1 1 1 1 1 1 1 1 1	N O N E

Legend

r = raw score = # of items answered correctly

np = # of possible unique response patterns yielding a given r. These values are based on the binomial expansion $(p + q)^n$, where n = 10.

Table 2: Expected Response Patterns at Nine Score Values for Normal and Uniform Ability Distributions

r	np	fN	fU	fN1	fU1	fN2	fU2
1	10	3	36	2	24	1	12
2	45	15	36	10	24	5	12
3	120	40	36	30	24	10	12
4	210	70	36	47	24	23	12
5	252	84	36	56	24	28	12
6	210	70	36	47	24	23	12
7	120	40	36	30	24	10	12
8	45	15	36	10	24	5	12
9	10	3	36	2	24	1	12

Legend

r = raw score = # of items answered correctly.

np = # of possible unique response patterns, given r.

fN = frequency of this score in a normal distribution.

fU = frequency of this score in a uniform distribution.

fN1 = frequency of Guttman response pattern for this score value in a normal distribution.

fU1 = frequency of Guttman response pattern for this score value in a uniform distribution.

fN2 = frequency of 2nd-most-likely response pattern in a normal distribution.

fU2 = frequency of 2nd-most-likely response pattern in a uniform distribution.

Table 3: Pattern Matrix Produced by Three Methods of Factor-Analyzing Normal Ability Distribution Data Set

	ϕ		r_{tet}		Rasch-r		Criterion	
ITEM	F1	F2	F1	F2	F1	F2	F1	F2
1	.125	.000	.039	.000	.505	.000	1.00	.000
2	.256	.000	.003	.000	.630	.000	1.00	.000
3	.385	.000	.391	.000	.729	.000	1.00	.000
4	.552	.000	.607	.000	.842	.000	1.00	.000
5	.609	.000	.808	.000	.891	.000	1.00	.000
6	.609	.000	.858	.000	.891	.000	1.00	.000
7	.552	.000	.858	.000	.841	.000	1.00	.000
8	.385	.000	.809	.000	.729	.000	1.00	.000
9	.256	.000	.603	.000	.629	.000	1.00	.000
10	.126	.000	.393	.000	.504	.000	1.00	.000
11	.000	.126	.000	.001	.000	.504	.000	1.00
12	.000	.256	.000	.000	.000	.629	.000	1.00
13	.000	.385	.000	.389	.000	.729	.000	1.00
14	.000	.552	.000	.603	.000	.842	.000	1.00
15	.000	.609	.000	.810	.000	.891	.000	1.00
16	.000	.609	.000	.858	.000	.891	.000	1.00
17	.000	.552	.000	.859	.000	.842	.000	1.00
18	.000	.385	.000	.810	.000	.729	.000	1.00
19	.000	.256	.000	.606	.000	.630	.000	1.00
20	.000	.126	.000	.393	.000	.505	.000	1.00

Table 4: Comparison of Three Methods of Factor-Analyzing Normal Ability Distribution Data Set

Method	ϕ	r_{tet}	Rasch-r	Criterion
r(*)	.7150	>1.000*	>1.000*	1.0000
$r_{max1}r_{max2}$	.3709	.7370	.7939	1.0000
$\Sigma\ h_i^2/20$	.1812	.3821	.5369	1.0000

Legend

r(*) = estimate of maximum internal correlation

$r_{max1}r_{max2}$ = product of two largest factor structure loadings

$\Sigma\ h_i^2/20$ = proportion of observed variance accounted for by the two common factors

Note: The presence of negative eigenvalues prevents the calculation of an exact r(*) value.

Table 5: Pattern Matrix Produced by Three Methods of Factor-Analyzing Uniform Ability Distribution Data Set

	ϕ		r_{tet}		Rasch-r		Criterion	
ITEM	F1	F2	F1	F2	F1	F2	F1	F2
1	.254	.000	.000	.000	.688	.000	1.00	.000
2	.434	.000	.494	.000	.769	.000	1.00	.000
3	.631	.000	.706	.000	.866	.000	1.00	.000
4	.758	.000	.898	.000	.931	.000	1.00	.000
5	.826	.000	.972	.000	.962	.000	1.00	.000
6	.826	.000	.996	.000	.962	.000	1.00	.000
7	.756	.000	.987	.000	.930	.000	1.00	.000
8	.629	.000	.937	.000	.865	.000	1.00	.000
9	.432	.000	.843	.000	.767	.000	1.00	.000
10	.252	.000	.715	.000	.687	.000	1.00	.000
11	.000	.253	.000	.071	.000	.688	.000	1.00
12	.000	.432	.000	.488	.000	.768	.000	1.00
13	.000	.631	.000	.701	.000	.866	.000	1.00
14	.000	.758	.000	.894	.000	.931	.000	1.00
15	.000	.826	.000	.972	.000	.963	.000	1.00
16	.000	.825	.000	.997	.000	.962	.000	1.00
17	.000	.758	.000	.985	.000	.930	.000	1.00
18	.000	.629	.000	.945	.000	.865	.000	1.00
19	.000	.433	.000	.844	.000	.767	.000	1.00
20	.000	.251	.000	.720	.000	.687	.000	1.00

Table 6: Comparison of Three Methods of Factor-Analyzing Uniform Ability Distribution Data Set

Method	ϕ	r_{tet}	Rasch-r	Criterion
r(*)	.8970	>1.000*	>1.000*	1.0000
$r_{max1}r_{max2}$	.6815	.9930	.9264	1.0000
Σ $hi^2/20$	.3807	.6562	.7205	1.0000

Legend

r(*) = estimate of maximum internal correlation

$r_{max1}r_{max2}$ = product of two largest factor structure loadings

Σ $hi^2/20$ = proportion of observed variance accounted for by the two common factors

Note: The presence of negative eigenvalues prevents the calculation of an exact r(*) value.

REFERENCES

Bentler, P.M. (1980). Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology, 31, 419-456.

Bock, R.D., & Lieberman, M. (1970). Fitting a response model for N dichotomously scored items. Psychometrika, 26, 347-372.

Christofferson, A. (1975). Factor analysis of dichotomized variables. Psychometrika, 40, 5-32.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.

Cureton, E.E., and D'Agostino, R.B. (1983). Factor analysis: An applied approach. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Davis, B.J., & Wright, B.D. (1988). MSCALE: A better way to measure. A Rasch measurement rating scale and test analysis computer program. Chicago: The University of Chicago, MESA Psychometric Laboratory.

Douglas, G.A., & Wright, B.D. (1990). Response patterns and their probabilities. Rasch Measurement SIG Newsletter, 3, No. 4, 2-4.

Ferguson, G.A. (1981). Statistical analysis in psychology and education. New York: McGraw-Hill Book Company.

Gerbing, D.W. (1989). A practical approach to structural equation modeling: When traditional item analysis becomes confirmatory factor analysis. Paper presented at the American Educational Research Assocation, San Francisco, CA.

Goldstein, H., & Wood, R. (1989). Five decades of item response modelling. British Journal of Mathematical and Statistical Psychology, 42, 139-167.

Guilford, J.P., & Fruchter, B. (1973). Fundamental statistics in psychology and education (5th ed.). New York: McGraw-Hill Book Company.

Harman, H.H. (1976). Modern factor analysis (3rd ed.). Chicago: The University of Chicago Press.

Hinkle, D.E., W. Wiersma, S.G. Jurs (1988). Applied Statistics for the Behavioral Sciences (2nd Ed.). Boston: Houghton Mifflin Company.

Joe, G.W., and Mendoza, J.L. (1989). The internal correlation: Its applications in statistics and psychometrics. Journal of Educational Statistics, 14, 211-226.

Jöreskog, K.G., & Sorböm, D. (1986). LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods (4th ed.). Mooresville, IN: Scientific Software, Inc.

Kachigan, S.K. (1986). Statistical analysis: An interdisciplinary introduction to univariate & multivariate methods. New York: Radius Press.

Kim, J.-O., & Mueller, C.W. (1978). Factor analysis: Statistical methods and practical issues. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-014. Beverly Hills and London: Sage Publications.

Kim, J.-O., Nie, N., & Verba, S. (1977). A note on factor analyzing dichotomous variables: the case of political participation. Political Methodology, 4, 39-62.

Lawley, D.N. (1943). On problems connected with item selection and test construction. Proceedings of the Royal Society of Edinburgh, 61, 273-287.

Loehlin, J.C. (1987). Latent variable models: An introduction to factor, path, and structural analysis. Hilldale, NJ: Lawrence Erlbaum Associates, Publishers.

Long, J.S. (1983). Confirmatory factor analysis. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-033. Beverly Hills and London: Sage Publications.

Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Mislevy, R.J. (1986). Recent developments in the factor analysis of categorical variables. Journal of Educational Statistics, 11, 3-31.

Mulaik, S.A. (1972). The foundations of factor analysis. New York: McGraw-Hill Book Company.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika, 49, 115-132.

Muthén, B.O. (1989). Dichotomous factor analysis of symptom data. Sociological Methods and Research, 18, 19-65.

Norusis, M.J. (1988). SPSS/PC+ V2.0 advanced statistics. Chicago: SPSS Inc.

Smith, R.M. (1988). The distributional properties of Rasch standardized residuals. Educational and Psychological Measurement, 48, 657-667.

SPSS, Inc. (1989). SPSS/PC+, V3.0. Chicago: SPSS, Inc.

Stage, F.K. (1988). University attrition: LISREL with logistic regression for the persistence criterion. Research in Higher Education, 29, 343-357.

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. Psychometrika, 52, 393-408.

Wise, S.L., & Tatsuoka, M.M. (1986). Assessing the dimensionality of dichotomous data using modified order analysis. Educational and Psychological Measurement, 46, 295-301.

Wright, B.D. (1977). Solving measurement problems with the Rasch model. Journal of Educational Measurement, 14, 97-116.

Wright, B.D., & Stone, M.H. (1979). Best test design. Chicago: MESA Press.